Efficiency of Monte Carlo Minimization Procedures and Their Use in Analysis of NMR Data Obtained from Flexible Peptides

HAGAI MEIROVITCH and EVA MEIROVITCH

Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306-4052

Received 12 February 1996; accepted 26 April 1996

ABSTRACT

The Monte Carlo minimization (MCM) method of Li and Scheraga is an efficient tool for generating low energy minimized structures of peptides, in particular the global energy minimum (GEM). In a recent article we proposed an enhancement to MCM, called the free energy Monte Carlo minimization (FMCM) procedure. With FMCM the conformational search is carried out with respect to the harmonic free energy, which approximates the free energy of the potential energy wells around the energy minimized structures (these wells are called localized microstates). In this work we apply both methods to the pentapeptide Leu-enkephalin described by the potential energy function ECEPP, and study their efficiency in identifying the GEM structure as well as the global harmonic free energy (GFM) structure. We also investigate the efficiency of these methods to generate localized microstates, which pertain to different energy and harmonic free energy intervals above the GEM and GFM, respectively. Such microstates constitute an important ingredient of our statistical mechanical methodology for analyzing nuclear magnetic resonance data of flexible peptides. Aspects of this methodology related to the stability properties of the localized microstates are examined.

Introduction

The biological function of proteins is determined to a large extent by their most stable 3-dimensional (3-D) native structure.[1,2] On the other hand, peptides and small organic molecules in solution generally populate several stable potential energy wells (which we call localized microstates[3–5]) in thermodynamic equilibrium.[6,7] It is important to develop efficient methods for a conformational search in order to predict these stable structures from theoretical considerations

CCC 0192-8651 / 97 / 020240-14

based solely on the interatomic interactions. This, however, is not an easy task because of the tremendous number of energy minima that exist on the energy surface of even a small peptide. Exact thermodynamic methods for a conformational search, such as the metropolis Monte Carlo (MC) method[8] and molecular dynamics[9] are very inefficient at room temperature, because the molecule typically becomes trapped in a low energy region close to the conformation from which the simulation started. Therefore, various techniques have been developed to surmount this problem; most of them, however, are at the expense of replacing the search for localized microstates of *low free energy* with a search for those of *low minimum energy*.[7, 10–15]

A very efficient procedure in the latter category is the MC minimization (MCM) method proposed by Li and Scheraga,[10] which applies the usual MC criterion to successively selected energy minimized structures i (of energy $E_i^{\min}$). To improve this method we proposed[3] to approximate the free energy of a localized microstate by the harmonic free energy, F_i^{har}, and to apply the metropolis criterion to F_i^{har} rather than $E_i^{\min}$. This procedure, called the free energy MCM (FMCM) method, has been applied so far to the pentapeptide Met-enkephalin described by the potential energy function ECEPP.[16] F_i^{har} consists of $E_i^{\min}$ and the harmonic entropy, which is obtained from the second derivatives of the energy with respect to the coordinates at the energy minimized structure. Therefore, if one uses a minimizer that is based on first and second analytical derivatives of the energy, FMCM requires only slightly more computer time than MCM.

The aim of the present article is to study the properties of MCM and FMCM as applied to models of Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu) described by ECEPP. Naturally we are interested in the efficiency of these methods to find the *global energy minimum* (GEM) structure, as well as the energy minimized structure that pertains to the localized microstate with the *global harmonic free energy minimum* (GFM). We also investigate the efficiency of MCM and FMCM to generate localized microstates in different energy and harmonic free energy intervals above the GEM and the GFM, respectively; as mentioned above, such distributions are important in theoretical studies of small molecules,[7] in particular for analyzing nuclear Overhauser effect (NOE) data obtained from multidimensional nuclear magnetic resonance (NMR) studies of peptides in solution. In this case the peptide can significantly populate several microstates, and the overall NOEs become averages over the contributions of the individual microstates, weighted by the populations of the latter.[6]

In previous studies we developed a methodology for analyzing such NMR data.[4, 5, 17] An important ingredient of this approach is the determination of larger microstates, called MC microstates, which are obtained with MC simulations starting from energy minimized structures (''seeds''). Typically an MC microstate contains a large number of localized microstates. In this context an important question is which seeds lead to more stable MC microstates; those with low energy, or those that pertain to localized microstates with low F^{har}. We shall try to answer this question as well.

Theory and Background

In this section we rigorously define the notion of a localized microstate and describe an exact MC method and two approximate MC procedures (MCM and FMCM) for identifying such microstates. We also provide a brief summary of our methodology for analyzing NMR data and will point out the importance of incorporating MCM and FMCM within its framework. As in previous work, the present calculations were carried out with ECEPP,[16] which is a relatively simple force field in which the dihedral angles are the only variables (see Results and Discussion). Note that we selected ECEPP for the sake of convenience; all the methods discussed in this study and our previous related work are applicable to any force field.

EXACT MC PROCEDURE FOR IDENTIFYING LOCALIZED MICROSTATES

Let us assume that a conformation of a peptide is defined by a set of internal variables ω that span the entire conformational space Ω, and that the energy $E(\omega)$ is given by a potential energy function such as ECEPP. We define a localized microstate Ω_i as the ensemble of conformations within the basin of attraction of conformation ω_i with *local* energy minimum E_i^{m}. Namely, a conformation ω pertains to microstate i if local energy minimization starting from ω leads to ω_i. The contribution Z_i of microstate i to the total partition function Z is

$$Z_i = \int_{\Omega_i} \exp[-E(\omega)/k_{\mathrm{B}}T]\, d\omega, \qquad (1)$$

where T is the absolute temperature and k_B is the Boltzmann constant. One can view Ω as a collection of microstates $\{\Omega_i\}$, where in the limit of a complete set of local energy minima $\Omega = \cup \Omega_i$, and Z can be obtained as $\sum Z_i$. This argument assumes a negligible contribution to Z from individual conformations corresponding to maxima and saddle points, which cannot naturally be assigned to a particular basin of attraction or inherent structure.[18a,b] (See also the application to protein dynamics in ref. 18c.) The free energy F_i of localized microstate i is

$$F_i = -k_B T \ln Z_i = E_i - TS_i, \quad (2)$$

where E_i is the Boltzmann average of the energy over Ω_i and S_i is the entropy.

In the Introduction we mentioned the trapping problem encountered with the usual MC method. To surmount this problem one may envisage an abstract *exact* MC procedure carried out among *localized microstates*, rather than *conformations*, as described below. Assume that the molecule resides at the localized microstate i; a trial microstate j is first selected with any transition probability T_{ij} that satisfied $T_{ij} = T_{ji}$. Then j is accepted with a probability p_{ij} that satisfies

$$p_{ij} = p_{ji} Z_j / Z_i = p_{ji} \exp\left[-(F_j - F_i)/k_B T\right]. \quad (3)$$

Assuming ergodicity, such a procedure ensures that the detailed balance condition is satisfied; therefore, microstate i will be distributed in direct proportion to Z_i. A major problem in implementing such a process lies in the difficulty to calculate Z_i at each MC step.

MCM PROCEDURE

A very useful approximation to the above exact procedure is provided by the MCM method developed by Li and Scheraga.[10] Let us describe this method as applied to a peptide described by ECEPP.[16] (Note, however, that MCM is not limited to rigid geometry force fields.[11c,19]) At step k of the process, the molecule is located at a local energy minimum E_i^m; a trial conformation is obtained by first selecting at random a dihedral angle (denoted ϕ_l) from the entire set of dihedral angles. Then a new value of ϕ_l is chosen uniformly from the range $[-180°$ to $180°]$, and the energy is minimized with respect to all the variables to define conformation j and T_{ij}; the latter is proportional to $\Delta_j \phi_l$, the range of ϕ_l values that lead to j by energy minimization. The probability p_{ij} of accepting j as the conformation of step $k + 1$ is

$$p_{ij} = \min\left(1, \exp\left[-(E_j^m - E_i^m)/k_B T^*\right]\right). \quad (4)$$

Contrary to eq. (3), where T denotes the thermodynamic temperature, T^* is a temperature parameter that can be optimized for efficiency; therefore, in this approximation Z_j/Z_i is replaced by $\exp[-(E_j^m - E_i^m)/k_B T^*]$. Notice, however, that the detailed balance condition is not satisfied, because $T_{ij} \sim \Delta_j \phi_l$ in general is different from $T_{ji} \sim \Delta_i \phi_l$. Therefore, the MCM procedure does not generate microstates with their Boltzmann probability Z_i. Nevertheless, the method has been found to be extremely efficient in locating low energy structures, in particular the GEM structure.[10,19–22] Li and Scheraga were able to find the GEM structure of the pentapeptide Met-enkephalin relatively easily using MCM[10] (see also refs. 7c and 19–22). However, one has to bear in mind that selecting a *single* angle per MC step probably does not allow for the structural rearrangement required for the process to be ergodic, especially for molecules larger than Met-enkephalin; therefore, the molecule may become trapped in a conformational region that does not include the GEM structure. Obviously, one can achieve ergodicity by generating a trial conformation that is based on changing *all* the angles. However, in this case the method would become inefficient because the trial structure is likely to be very different from the existing one, and the benefit of the gradual structural optimization characterizing the MC process would be lost.

Indeed, in a later article[10b] Li and Scheraga report that the GEM structure can be located with better efficiency using a procedure in which m trial angles are selected with probability 2^{-m}, whereby a single angle is changed 50% of the time, two angles 25% of the time, etc. It should be pointed out, however, that the efficiency of locating the GEM structure, even with an ergodic MCM procedure, depends to a large extent on the complexity of this structure. For example, one would expect to be able to generate relatively easily an α-helical GEM structure, while a structure that consists of knots might in practice be impossible to obtain. To further improve the efficiency of MCM, the temperature T^* was changed in various thermalization schedules.[19,21]

FMCM PROCEDURE

The FMCM procedure introduced by Vásquez et al.[3] was conceived as an improvement over the MCM method, by replacing the energy in eq. (4) by the harmonic free energy F_i^{har}. Thus, with FMCM a trial conformation j is obtained as before. However, at both energy minima i and j one now also calculates the Hessian, which is the matrix of second derivatives of the energy with respect to the dihedral angles. This enables estimation of the entropy S_i^{har} within the harmonic approximation, which up to an additive constant is[23–25]

$$S_i^{\mathrm{har}} = -1/2 \ln[\det(\mathrm{Hessian})], \qquad (5)$$

where det stands for determinant. The harmonic approximation of the free energy is thus

$$F_i^{\mathrm{har}} = E_i^m - TS_i^{\mathrm{har}}. \qquad (6)$$

T is the thermodynamic temperature, which should be distinguished from the efficiency parameter T^* [eq. (4)]. It should be pointed out that the average energy includes, besides E_i^{m}, the term $Nk_{\mathrm{B}}T/2$, where N is the number of variables. This term is the same for all the localized microstates and has therefore been omitted from eq. (6). We shall refer to E_i^{m} as the energy of localized microstate i. Also, notice that, while S_i^{har} is independent of T, at high T it dominates F_i^{har} through the term TS_i^{har}. The harmonic partition function is

$$Z_i^{\mathrm{har}} = \exp\left[-F_i^{\mathrm{har}}/k_{\mathrm{B}}T\right]. \qquad (7)$$

The metropolis criterion is now applied by replacing E_i^{m} and E_j^{m} in eq. (4) by F_i^{har} and F_j^{har}, respectively. Thus,

$$p_{ij} = \min\left\{1, \exp\left[-\left(F_j^{\mathrm{har}} - F_i^{\mathrm{har}}\right)/k_{\mathrm{B}}T^*\right]\right\}. \qquad (8)$$

The advantage of this procedure is obviously in the inclusion (even though approximately) of entropic contributions.

With FMCM localized microstates are not distributed according to their Boltzmann probability ($\sim Z_i$) or the harmonic probability ($\sim Z_i^{\mathrm{har}}$) even for $T = T^*$. However, if a large enough set of localized microstates of low F_i^{har} is produced, statistical averages in the harmonic approximation can be calculated using the known populations that are proportional to Z_i^{har}. Notice that such averages can be obtained also with MCM because one can calculate F_i^{har} for each of the energy minimized structures obtained in the simulation; this indeed is done in the present work.

At low temperatures, where the energy dominates the free energy, the GEM and GFM structures are expected to be the same, or very close to one another, and therefore should be identified with equal efficiency by MCM and FMCM. This picture changes at high temperature, where TS_i^{har} dominates F_i^{har}, implying that S_i^{har}(GFM) will be larger than S_i^{har}(GEM); one would also expect E_i^{m}(GFM) to be larger than the GEM because we previously found that, on average, localized microstates with higher harmonic entropy also have higher energy.[4] It should be noted that E_i^{m} is correlated with its structure ω_i in the sense that as E_i^{m} decreases, ω_i acquires more of the conformational characteristics of the GEM structure. Therefore, a search for the GEM structure with a procedure that depends on the energy, such as MCM, is relatively efficient. On the other hand, the entropy, which reflects the local fluctuations around the energy minimized structure, is less correlated with ω_i. Therefore, at high temperature, where S_i^{har} dominates F_i^{har}, the search for the GFM structure is expected to be less efficient than the search for the GEM structure, not only with MCM, but also with FMCM.

STABLE MICROSTATES AND ANALYSIS OF NMR DATA OF FLEXIBLE PEPTIDES

As pointed out in the Introduction, we are not interested only in the GEM and GFM structures, but also in the efficiency of MCM and FMCM to generate a larger group of highly stable localized microstates, such as is required to model flexible molecules. For example, a linear peptide in solution, while relatively flexible, can still produce medium- and long-range NOEs, inferring that it does not prevail as a random coil, but significantly populates several microstates in thermodynamic equilibrium. Such *intermediate* flexibility can also be found in loops and side chains of proteins. Determination of the most populated states of a flexible molecule from multidimensional NMR data is still an open question,[6] which motivated us to develop for that purpose a new methodology based on statistical mechanical principles.[4, 5, 17] Thus far we have treated Leu-enkephalin described by ECEPP, as well as by ECEPP with additional terms that take into account solvation effects in an implicit way.[17]

In the first stage of this methodology a large number of conformations is generated *at random* and their energy is minimized. One then selects a large energy range (say, 12 kcal/mol) above (and including) the GEM and divides it into small intervals (bins) j (we chose 0.5 kcal/mol). A localized microstate i is assigned to interval j if its minimum energy E_i^{m} falls within the interval limits (notice that the microstates that pertain to the 12 kcal/mol range provide most of the contribution to Z). A good approximation for Z_i [eq. (1)] was obtained by an average based on Z_i^{har} [eq. (7)] and another approximation for Z_i obtained with the ''counting method'' (see ref. 4). The Z_i values of the localized microstates that belong to bin j are added and their sum is divided by the combined contribution of all the bins; this ratio constitutes an estimate for Z_j/Z. We found that for Leu-enkephalin the energy range of 2 kcal/mol above the GEM contributes 0.6 of Z, and we adopted this range as the significant one; i.e., its microstates should reproduce the experimental results. Because the random search discussed above is relatively inefficient, in order to identify the maximum number of localized microstates we carried out an extensive conformational search in the 2 kcal/mol range by a procedure called SADA (systematic alteration of dihedral angles).[26] SADA is based on systematic and stochastic search parts. In the systematic part the dihedral angles are changed sequentially, one at a time, in increments of n_{grid} degrees within the range $[-180°, 180°]$, and the energy is calculated. If a lower energy is obtained for an altered dihedral angle, the new values of the energy and the angle replace the current ones; if not, the current values are retained. Obviously, the conformational search could have been carried out by any other technique, in particular MCM or FMCM.

It should be pointed out that already within the 2 kcal/mol range the number of energy minimized structures is large, and many of them are similar. To avoid this redundancy and the need to identify all the localized microstates (an untractable task for larger molecules), we defined the so-called MC microstate (to be distinguished from the localized microstate), which is a broader region of conformational space that typically includes many localized microstates. More specifically, this is the region spanned by an MC simulation that starts from a minimum energy structure (''seed'') and is characterized by a relatively large dihedral angle variability. The MC microstates should cover the region defined by the localized microstates. However, they are not allowed to overlap, and therefore the MC simulations should start from a set of seeds that are *significantly* different. A criterion for difference between two structures adopted in previous studies is that at least one angle differs by 60° or more. The free energies of the MC microstates (hence their populations) are then obtained by the local states method.[27] With this method the range of change in the sample of each dihedral angle is calculated; each range is divided into an equal number of segments and transition probabilities between segments of successive angles along the chain are obtained. This enables one to express approximately the probability of a conformation as a product of the corresponding transition probabilities, which leads to an approximation of the entropy. Finally, the individual contributions of the MC microstates to the NOEs are calculated, and the overall NOE intensity (which is compared to the experimental result) is an average of these contributions weighted by the populations of the MC microstates.

The above classification of the localized microstates into bins according to the energy E_i^{m} is not unique. In a previous study we also ranked them according to the harmonic free energy, studied their distribution in bins of 0.5 kcal/mol above the GFM, and investigated the relative contribution of these bins to the harmonic partition function.[4] Within the 2 kcal/mol range the two classifications led to comparable relative contribution to the partition function.

However, within the framework of our methodology, the more effective classification is the one that leads more effectively to the most stable MC microstates. Note that the low harmonic free energy bins (unlike the corresponding energy bins) also contain localized microstates with relatively high energy, but large harmonic entropy. The question is whether this large entropy is only a local property, or whether it also correlates with a large entropy of the corresponding MC microstate. The answer to this question is not clear, and will be discussed later.

Results and Discussion

MODELS

As mentioned previously, the properties of MCM and FMCM were investigated as applied to Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH). The

molecule was modeled by the potential energy function ECEPP/2, which assumes rigid geometry (i.e., constant bond lengths and angles), and was based on nonbonded, electrostatic, torsional, and hydrogen-bond potentials.[16] Two models of ECEPP were studied: in one (called model I) the peptide bond angles ω were kept fixed at 180°; therefore, a conformation was defined solely by 19 variables, the 10 backbone dihedral angles ϕ and ψ and the nine side chain dihedral angles χ. With the second model (called model II) the five backbone ω angles were allowed to vary; therefore, a conformation was defined by 24 variables. In previous studies we investigated Leu-enkephalin at $T = 280$ K; for comparison we simulated this molecule here at the same temperature and also at 400 K. We used the standard dielectric constant $\varepsilon = 2$ of ECEPP.

MCM AND FMCM SIMULATIONS

The simulations were carried out with the program FANTOM,[21,28] which is based on the ECEPP potential. FANTOM enables the minimization of the energy with the Newton–Raphson method, and provides the first and second analytical derivatives of the energy with respect to the dihedral angles. Thus, the calculation of the Hessian determinant and the harmonic entropy [eq. (5)], which leads to the free energy [eq. (6)], is straightforward. To obtain reliable results for the free energy, the energy minima should be obtained with high precision. We ensured this by setting an absolute value of 10^{-9} and 10^{-10} kcal/mol/deg for the gradient as tolerance for stopping the minimization. If the matrix was found to be not positive definite, the corresponding energy minimum was discarded and another one was generated; this happened in less than 1 and 1.5% of the MC steps for models I and II, respectively.

In an attempt to optimize MCM and FMCM, we tested several MC protocols (including the ''shock'' and ''adaptive'' schedules provided by FANTOM); for Leu-enkephalin they all were found to be equally efficient, with no clear advantage of one over the others. With the procedure used in this study, a trial conformation was determined in several steps. First one determines m, the number of trial angles to be changed, as

$$m = \min[N, \mathrm{int}(1. - \ln(p + 0.00001))], \quad (9)$$

where N is the number of variables, p is a random number within the range $[0, 1]$, and $\mathrm{int}(a)$ is the integer value closest to a from below. Thus, the probability for $m = 1$ and 2 is ~ 0.63 and 0.24, respectively. The specific angles m are then determined at random, where side chain angles are selected with a relatively low probability of 0.06. Each selected angle is then changed at random within the range $\pm 0.85\pi$ around its current value.[20] Note that even with model II, which allows the angles ω to vary, only the dihedral angles ϕ, ψ, and χ (but not the ω angles) are changed at this stage. However, the following energy minimization is applied to all the angles including ω. A similar procedure was adopted in most of the previous studies of enkephalins by MCM.[10,21] With both MCM and FMCM, if the energies of the trial and the present conformations differ by less than 0.00001 kcal/mol, the trial conformation is rejected.

We define the acceptance rate as the number of MC steps accepted divided by t_{tot}, the total number of MC steps carried out in the simulation. For each set of n simulations $n/2$ were carried out using $T^* = 300$ K and the other half using $T^* = 600$ K; the efficiency for identifying the GEM and GFM structures was found to be the same for both temperatures. As expected, the acceptance rate was larger for the higher temperature; for example, for model II (variable ω) simulated by MCM it was ~ 0.09 and ~ 0.18 at $T^* = 300$ and 600 K, respectively. Similar behavior was observed for the FMCM simulations, where the corresponding results for the same model at $T = 280$ K were ~ 0.11 and ~ 0.24. The slightly larger value obtained with FMCM stems from the fact that on average the difference between the energies of two localized microstates is larger than the difference between their harmonic free energies, as will be discussed in a later section. Because we are also interested in the performance of MCM and FMCM in generating different structures, we stored in a file, not only the conformations that were accepted by the MC criterion, but also those that were rejected. Only different conformations were retained, where the criterion for variance for two conformations is that at least one angle differs by more than 1°; the symmetries of χ_2(Phe) χ_3(Leu) and χ_4(Leu) were taken into account. For each conformation we also stored in the file the energy E_i^{m} and the harmonic entropy S_i^{har} [eq. (5)], which was also calculated in the MCM simulations. It should be pointed out that for both MCM and FMCM, only conformations with energy smaller than 2 kcal/mol were stored; thus, an overall energy range of approximately 12 kcal/mol above the GEM was considered in these calculations (see Table I).

TABLE I.
Dihedral Angles of Global Energy Minimum (GEM) and Global Harmonic Free Energy (GFM) Structures.

		Model I ($\omega = 180°$)			Model II (Variable ω)		
Residue	Angle	GEM	GFM $T = 280$ K	GFM $T = 400$ K	GEM	GFM $T = 280$ K	GFM $T = 400$ K
Tyr[1]	ϕ	−86.7	−86.7	−84.4	−78.2	−86.5	−87.6
	ψ	152.9	153.7	144.5	136.8	153.0	156.3
	ω	180.0	180.0	180.0	−168.9	−178.9	−179.9
	χ^1	−179.5	−179.6	−177.3	−165.0	179.9	−178.9
	χ^2	70.3	68.5	76.6	70.7	67.0	75.8
		(−109.7)	(−111.5)	(−103.4)	(−109.3)	(−113.0)	(−104.2)
	χ^6	−35.1	−34.8	**−9.8**	−159.6	−33.7	−179.5
		(144.9)	(145.2)	(**170.2**)	(20.4)	(146.3)	(0.5)
Gly[2]	ϕ	−161.1	−161.5	−158.0	−87.4	**−160.2**	**−178.1**
	ψ	72.1	70.6	56.7	65.5	67.5	69.6
	ω	180.0	180.0	180.0	176.1	177.7	179.2
Gly[3]	ϕ	64.4	64.4	**98.9**	65.7	67.4	66.4
	ψ	−93.0	−93.6	**−16.4**	−92.7	−93.6	−95.4
	ω	180.0	180.0	180.0	177.4	177.0	176.1
Phe[4]	ϕ	−83.9	−82.8	**−156.6**	−83.5	−81.3	−69.9
	ψ	−25.0	−26.2	**159.3**	−33.5	−28.3	−33.5
	ω	180.0	180.0	180.0	177.4	−178.7	175.8
	χ^1	72.7	**−179.8**	58.2	178.7	179.7	179.4
	χ^2	−95.4	74.6	−88.2	77.3	74.5	−99.6
Leu[5]	ϕ	−78.6	−82.0	−80.0	−155.5	**−80.8**	**−73.3**
	ψ	130.5	141.8	121.0	−75.9	**142.9**	**135.0**
	ω	180.0	180.0	180.0	−177.8	−178.7	−178.5
	χ^1	179.7	**−56.9**	180.0	178.1	**−57.6**	176.9
	χ^2	65.6	**170.0**	63.9	61.9	**168.7**	66.3
	χ^3	−172.9	−59.4	−67.9	−68.1	60.4	174.3
	χ^4	60.6	−171.9	179.3	59.1	67.6	59.8
E_i^m		−9.704	−9.464	−4.885	−10.093	−9.611	−7.676
F_i^{har} (280 K)		9.982	9.196	9.778	15.538	14.434	14.558
F_i^{har} (400 K)		18.417	17.193	**16.062**	26.523	24.739	**24.087**

Each energy (and free energy) is obtained for two structures with different combination of χ^2 and χ^6 of Tyr (one combination appears in parentheses in the table), while the remaining angles differ by less than 1°. For each model the angles of the GFM structures that differ from the corresponding angles of the GEM structure by more than 20° are boldfaced. The minimum energy E_i^m and the harmonic free energy F_i^{har} of the corresponding localized microstates at 280 and 400 K are presented in the rows at the bottom. The GFM values at 400 K are boldfaced.

For both models the simulations were started from randomly chosen conformations and were carried out with different sequences of random numbers. Three sets of simulations were performed, two by FMCM at $T_{MC} = 280$ and 400 K, and one by MCM. For model I ($\omega = 180°$) a set consists of $n = 14$ simulations, each of $t_{tot} = 10^4$ MC steps; for model II, $n = 10$ and $t_{tot} = 2 \times 10^4$. It should be pointed out that because the energy and harmonic entropy are known for each structure, one can calculate the corresponding harmonic free energy at any given temperature. Therefore, sometimes we distinguish between two kinds of T [eq. (6)]: T_{MC}, the temperature at which F_i^{har} [eq. (6)] is calculated during the FMCM simulations, and T_{ana}, the temperature used to calculate the values of F_i^{har} of a given sample (generated with FMCM at any temperature, or with MCM) in the analysis stage. Thus, altogether three different temperatures were involved in this study: T^* [eqs. (4) and (8)], which is an efficiency parameter, and the thermodynamic temperatures T_{MC} and T_{ana}, which are both related to T, as defined in eq. (6). A simulation $t_{tot} = 10^4$ MC steps of model I with a gradient tolerance of 10^{-10} requires ~ 4 h CPU time on the IBM 9076 SP2 RISC processor.

GENERATING GEM AND GFM STRUCTURES

We first investigated the efficiency of MCM and FMCM to generate the GEM structure, and the GFM structures at $T = 280$ and 400 K. These structures, and the corresponding energies and harmonic free energies at $T = 280$ and 400 K, appear in Table I. It should be pointed out that within a precision of 0.001 kcal/mol, each structure in this table has a twofold degeneracy; i.e., the same energy and harmonic free energy are obtained for two combinations of the values of χ^2 and χ^6 of Tyr (they appear in parentheses in Table I). The remaining angles are basically the same; i.e., they differ by less than 1°. We considered both members of the pair as the GEM (or GFM) structure. The results for efficiency appear in Table II. For each sample the program finds the MC steps $t^{\mathrm{m}}_{\mathrm{GEM}}$ and $t^{\mathrm{m}}_{\mathrm{GFM}}$ at which the corresponding GEM and GFM structures are obtained for the first time; their minimum values, $t^{\mathrm{min}}_{\mathrm{GEM}}$ and $t^{\mathrm{min}}_{\mathrm{GEF}}$, and their averages, $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$ and $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$, over each set are provided in Table II. These averages were calculated only for the n_{u} samples for which the GEM or GFM structures were obtained.

We now discuss the results shown in Table II for model I ($\omega = 180°$) obtained with MCM ($T_{\mathrm{ana}} = 280$ K) and FMCM ($T_{\mathrm{ana}} = T_{\mathrm{MC}} = 280$ K). With both methods $n_{\mathrm{u}} = 12$; i.e., in two out of the 14 simulations the GEM and GFM structures were not reached. This means that the averages $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$ and $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$ were actually larger than the values provided. This fact, and the large fluctuations observed in the values of $t^{\mathrm{m}}_{\mathrm{GEM}}$ and $t^{\mathrm{m}}_{\mathrm{GFM}}$ (e.g., ranging from 66 to a value larger than 10^4 for the GFM) suggest that the use of the above average values as a measure of efficiency should be adopted with caution. To enhance the reliability of this measure, both n and the sample size t_{tot} should be increased.

As one would expect, $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$ was smaller (even though only slightly) than $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$ for MCM, but $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}} > \bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$ for FMCM. Indeed, with MCM the GEM structure appeared prior to the GFM one in seven of the 12 samples; with FMCM the GFM structure was encountered first in 10 of the 12 samples. Also, the expected relations $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}(\mathrm{MCM}) < \bar{t}^{\mathrm{m}}_{\mathrm{GEM}}(\mathrm{FMCM})$ and $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}(\mathrm{MCM}) > \bar{t}^{\mathrm{m}}_{\mathrm{GFM}}(\mathrm{FMCM})$ were satisfied. This means that for this model MCM is a slightly better tool than FMCM for identifying the GEM structure, while FMCM is the more efficient procedure for locating the GFM conformation. However, the efficiencies of the two methods are very similar; this is a consequence of the similarity between the GEM and GFM structures (they differ significantly only in the values of χ^1 of Phe and Leu and χ^2 of Leu, which are boldfaced in Table I) and their close energies and harmonic free energies, −9.704 vs. −9.464, and 9.982 vs. 9.196 kcal/mol, respectively.

It is of interest to compare these results to those obtained for model II (variable ω) with MCM ($T_{\mathrm{ana}} = 280$ K) and FMCM ($T_{\mathrm{ana}} = T_{\mathrm{MC}} = 280$ K). As expected, the averages $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$ and $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$ for this model are larger (due to a more complex confor-

TABLE II.
Efficiency of MCM and FMCM to Locate GEM and GFM Structures.

	T_{MC}	T_{ana}	n	t_{tot}	$\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$	$t^{\mathrm{min}}_{\mathrm{GEM}}$	n_{u}	$\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$	$t^{\mathrm{min}}_{\mathrm{GFM}}$	n_{u}
Model I ($\omega = 180°$)										
MCM		280	14	10^4	2700 ± 2500	304	12	3400 ± 2800	139	12
FMCM	280	280	14	10^4	3700 ± 3400	211	12	2400 ± 2700	66	12
FMCM	400	400	14	10^4	3300 ± 3500	399	11	2300 ± 2200	823	3
Model II (variable ω)										
MCM		280	10	2×10^4	5000 ± 4400	907	9	3700 ± 3200	772	10
FMCM	280	280	10	2×10^4	8700 ± 4200	2100	9	3900 ± 3100	186	10
MCM		400	10	2×10^4				4500 ± 3700	329	9
FMCM	280	400	10	2×10^4				9300 ± 6300	1148	9
FMCM	400	400	10	2×10^4	6800 ± 5000	1280	7	5000 ± 3600	531	7

T_{MC} is the temperature (in degrees K) at which the FMCM simulations were carried out. T_{ana} is the temperature at which the harmonic free energy F_i^{har} [eq. (6)] of each member of the sample was calculated in the analysis stage from the available energy E_i^{m} and the harmonic entropy S_i^{har} [eq. (5)]. n is the total number of simulations in a set, n_{u} is the number of simulations in which the GEM or the GFM structure were obtained, and t_{tot} is the total number of MC steps in a simulation. $\bar{t}^{\mathrm{m}}_{\mathrm{GEM}}$ and $\bar{t}^{\mathrm{m}}_{\mathrm{GFM}}$ are the average (over the n_{u} samples) MC step at which the GEM and GFM structures, respectively, appeared for the first time. $t^{\mathrm{min}}_{\mathrm{GEM}}$ ($t^{\mathrm{min}}_{\mathrm{GFM}}$) is the smallest MC steps at which the GEM (GFM) structure appeared for the first time in the n_{u} samples.

mational space) than the corresponding values obtained for model I. (This is why the sample size t_{tot} was doubled for model II.) As for model I, the GEM structure was generated more efficiently by MCM than by FMCM ($\bar{t}^{m}_{GEM}$ = 5000 vs. 8700), and with FMCM the GFM structure was reached before the GEM structure (in nine of the 10 samples). For the two methods $\bar{t}^{m}_{GFM}$ was also comparable. However, unlike for model I, MCM found the GFM structure more efficiently than the GEM structure. (The former was obtained prior to the latter in eight of the 10 samples.) This probably stems from the fact that the GFM structure pertains to a region in conformational space that is less isolated than the GEM region and is therefore reached easily by the MCM simulation. Indeed, these structures differ significantly (in four backbone and three side chain angles, which are boldfaced in Table I), even though their energies are relatively close (-10.09 vs. -9.61 kcal/mol). This view is also supported by the fact that the GFM structure was obtained in *all* the samples (n_u = 10), while the GEM structure occurred only in $n_u = 9$ of the samples obtained with both methods.

Because at $T = 280$ K F_i^{har} of both models is dominated by the energy, the GEM and the energy of the GFM structure are close (see Table I) and the difference in efficiency between MCM and FMCM is not large. Therefore, we also simulated the two modeled with FMCM at $T_{MC} = 400$ K, where the contribution of the entropy to F_i^{har} becomes more significant. Indeed, for model I ($\omega = 180°$) the energy of the GFM structure at $T = 400$ K (-4.88 kcal/mol) is ~ 5 kcal/mol above the GEM (-9.70 kcal/mol), while a smaller difference of ~ 2.4 kcal/mol (-10.09 vs. -7.68 kcal/mol, see Table I) was observed for model II. Table I shows that for model I the GFM structure was reached only in three of the 14 FMCM samples simulated at T_{MC} = 400 K.

This, however, is still a better efficiency than that obtained with MCM; thus, we searched the MCM samples using $T_{ana} = 400$ K and found the GFM in a single sample only. On the other hand, the GEM structure was identified quite efficiently even with FMCM at $T_{MC} = 400$ K; i.e., in 11 of the 14 samples. These results can be explained in the same way as before: in addition to their substantially different energies, the GEM and GFM structures at 400 K differ significantly from each other (by the five backbone and two side chain angles, which are boldfaced in Table I). The GFM structure probably belongs to a region in conformational space that has a limited accessibility with the MCM simulations, even though the latter typically span a relatively large energy range of ~ 15 kcal/mol. This structure is not reached easily by FMCM because at $T = 400$ K the free energy is dominated to a large extent by the entropy, which is due to its local character not providing a strong guidance for finding the GFM structure (see discussion in Theory and Background).

The picture is different for model II (variable ω). First it should be pointed out that the GFM (400 K) structure is very similar to the GFM (280 K) structure. (They differ significantly only by χ^1 and χ^2 of Leu, see Table I.) Because the GFM (280 K) structure appears in all the FMCM samples simulated at 280 K, as well as all the MCM samples, one would also expect to find the GFM (400 K) structure in most of these samples; indeed, it appears in nine out of 10 samples (see Table II). These simulations seem to be more efficient than the FMCM simulations at $T_{MC} = 400$ K, where the GFM (400 K) structure is found only in seven samples. This, however, might reflect a statistical fluctuation because in two of the other three FMCM samples, structures with harmonic free energy that exceeds the GFM (400 K) by only ~ 0.1 kcal/mol were obtained.

In summary, the calculations for Leu-enkephalin presented thus far do not demonstrate a large difference in the efficiencies of the two methods. However, in general MCM appears to be the more efficient procedure for generating the GEM structure, while FMCM is found to be the better procedure for generating the GFM structure. One would expect these conclusions to hold for other molecules as well, even though exceptions can occur (as demonstrated above) because the performance of both methods depends on the energy surface of the molecule (see discussion in Theory and Background). For larger molecules the probability for $m > 1$ [eq. (9)] should be increased and thermalization schedules applied.[19,21]

GENERATION OF LOW ENERGY AND FREE ENERGY LOCALIZED MICROSTATES

As previously,[4] we also studied the efficiency of the two methods to generate low energy and low harmonic free energy localized microstates as the energy and the harmonic free energy are increased above the GEM and GFM, respectively. This was carried out by analyzing the six sets of samples mentioned above. First we combined the n files of each set to a larger file of l_{tot} structures

(see Table III). Then from each of these combined files we obtained a smaller file of l_{50} structures that were *significantly* different; i.e., for every pair of conformations at least one angle differed by 50° or more. This criterion is somewhat less strict than the one we used earlier,[5] which was based on 60° and was not applied to ϕ(Tyr), χ^6(Tyr), and ψ(Leu). From the latter samples we calculated the number of occurrences of the significantly different structures with harmonic free energy and energy bins of 0.5 kcal/mol above the GFM (Table III) and GEM (Table IV), respectively. It should be pointed out that these numbers depend slightly on the order of the l_{50} structures in the file. Also, notice that similar calculations were carried out in ref. 4; however, in this work the conformations were generated at random.

Table III reveals that the number of structures found with both methods is approximately equal for only the first two or three bins (boldfaced results in Table III). For the higher free energy bins FMCM generates a significantly larger number of structures than MCM, and the difference increases as the free energy of the bin increases. The similar results obtained for the first few bins are due to the fact that these bins are the least populated, and probably all their members already exist in the relatively large samples generated with both methods. Obviously for extremely large samples one would expect to obtain the same number of structures also in higher free energy bins. However, for samples of practical size, FMCM is clearly more efficient than MCM.

In an earlier study[4] we found that on average the localized microstates of higher energy also have larger harmonic entropy. Therefore, the bins of low harmonic free energy are also populated by localized microstates of relatively high energy, which gain their stability from the large entropy. This is demonstrated in Table V, where for each free energy bin (at $T = 280$ K) the minimum and maximum energies obtained in the simulations, and their difference are provided. Thus, in the 2 kcal/mol range, these differences are 6.53 and 6.12 kcal/mol for models I and II, respectively; and they increase for the larger free energy bins. Similar results were obtained at $T = 400$ K (not shown) where, as expected, the corresponding differences were larger. Note that Table III does not provide the populations of harmonic free energy bins larger than 6 kcal/mol; this is because we found that, due to the entropy increase, such bins should

TABLE III.
Distribution of Localized Microstates According to Their Harmonic Free Energy F_i^{har} [Eq. (6)] in Bins of 0.5 kcal / mol Above GFM.

	Model I				Model II			
	$T = 280$ K		$T = 400$ K		$T = 280$ K		$T = 400$ K	
Bin	MCM	FMCM	MCM	FMCM	MCM	FMCM	MCM	FMCM
0.0–0.5	**6**	**6**	**4**	**4**	**8**	**8**	**10**	**12**
0.5–1.0	**18**	**17**	**0**	**0**	**32**	**33**	38	57
1.0–1.5	**43**	**44**	18	28	**95**	**97**	105	153
1.5–2.0	84	94	42	52	267	300	211	309
2.0–2.5	288	332	123	171	440	515	461	670
2.5–3.0	435	517	295	453	699	825	644	1122
3.0–3.5	808	1081	549	1347	1031	1355	1065	1825
3.5–4.0	1138	1491	851	1716	1475	1874	1468	2395
4.0–4.5	1548	1957			1911	2519		
4.5–5.0	1937	2392			2129	2826		
5.0–5.5	2223	2727			2615	3455		
5.5–6.0	2511	3086						
l_{tot}	60077	63607	60077	61041	93766	95490	93766	93995
l_{50}	30129	32804	30144	33029	34133	40128	34127	40661

The energy minimized structures of these microstates are significantly different; i.e;, for every two structures at least one angle differs by 50° or more. The results were obtained from the six sets of simulations described in Table II. The total number of structures in the set of n samples is l_{tot} (to be distinguished from the total number of MC steps). Only l_{50} of these structures are significantly different. The FMCM samples were generated at the given temperatures, whereas the MCM samples were analyzed at these temperatures (i.e., for each structure the free energy was calculated from the values of E_i^m and S_i^{har}). For each pair of columns (based on MCM and FMCM) the first several bins are equally populated; these numbers are boldfaced.

TABLE IV.
Number of Energy Minimized Structures Significantly Different in Energy Bins of 0.5 kcal / mol Above GEM Obtained by MCM and FMCM.

	Model I			Model II		
Bin	MCM	FMCM 280 K	FMCM 400 K	MCM	FMCM 280 K	FMCM 400 K
0.0–0.5	**8**	**8**	**8**	**12**	**12**	**13**
0.5–1.0	**10**	**10**	**10**	**38**	**37**	29
1.0–1.5	**16**	**15**	**17**	46	39	31
1.5–2.0	**54**	**53**	44	91	82	58
2.0–2.5	100	91	74	181	159	122
2.5–3.0	207	189	68	315	249	189
3.0–3.5	284	242	212	435	353	286
3.5–4.0	468	397	322	675	548	454
4.0–4.5	626	538	440	853	806	677
4.5–5.0	1050	916	770	1126	1032	813
5.0–5.5	1345	1169	1029	1510	1454	1243
5.5–6.0	1639	1599	1474	1723	**1852**	1718
6.0–6.5	1858	**1907**	1719	1955	2199	**1965**
6.5–7.0	2041	2148	1975			2351
7.0–7.5	2307	2583	**2474**			

See caption of Table III. The total number of significantly different structures, I_{50} appear in Table III. For each model the first several bins are equally populated; these numbers are boldfaced (see discussion in the text). For the higher energy bins the MCM values are larger than the corresponding FMCM (280 K) values, and the latter are larger than the FMCM (400 K) results. These relations change at $\sim$ 6 kcal, where the values of FMCM (280 K) (boldfaced) exceed those of MCM. Also boldfaced are the results for FMCM (400 K), which are larger than their MCM counterparts. At $\sim$ 8.5 kcal / mol the FMCM (400 K) values become larger than the FMCM (280 K) ones (these values do not appear in the table).

contain conformations with $E_i^{\mathrm{m}} > 2$ that are not included in our samples. This simultaneous increase in both energy and entropy also means that a relatively large number of energy–entropy combinations can lead to low harmonic free energy; therefore, the free energy bins are expected and, indeed, found to be more populated than the corresponding *energy* bins of Table IV. Therefore, on average, the difference in the harmonic free energy between a pair of structures that are compared

TABLE V.
Results for Distribution of Energy in *Harmonic* Free Energy Bins *j* of 0.5 kcal / mol Above GFM at *T* = 280 K.

	Model I			Model II		
Bin(j)	$E_j^{\min}$	$E_j^{\max}$	ΔE_j	$E_j^{\min}$	$E_j^{\max}$	ΔE_j
0.0–0.5	0.24	2.48	2.34	0.48	2.42	1.94
0.5–1.0	0	4.82	4.82	0.82	3.07	2.25
1.0–1.5	0.25	4.82	4.57	0	5.15	5.15
1.5–2.0	0.77	6.53	5.76	0.54	6.12	5.58
2.0–2.5	1.14	6.61	5.47	0.27	6.68	6.41
2.5–3.0	1.17	7.38	6.21	0.42	8.38	7.96
3.0–3.5	1.54	8.44	6.90	0.85	9.15	8.30
3.5–4.0	1.66	8.92	7.26	1.16	9.47	8.31
4.0–4.5	1.79	9.91	8.12	2.01	10.32	8.31

For each bin the lowest and highest minimized *energies* were found and their differences $E_j^{\min}$ and $E_j^{\max}$ from the GEM (in kcal / mol) were calculated. $\Delta E_j = E_j^{\max} - E_j^{\min}$ is also provided. It is evident that ΔE_j is always much larger that the bin size, 0.5 kcal / mol, and that it increases for the higher free energy bins.

during the FMCM simulation is smaller than the corresponding energy difference obtained with MCM, indicating that the acceptance rate is larger for the former than for the latter, as already shown. The fact that high energy microstates populate the low free energy bins also suggests that a larger diversity of structures is obtained with FMCM than with MCM. This is, indeed, supported by the results for l_{50}, presented in the row at the bottom of Table III: for each pair, l_{50} is always larger for FMCM than for MCM.

As mentioned previously, Table IV presents the number of significantly different conformations in energy bins of 0.5 kcal/mol above the GEM; in this classification the localized microstates are defined by their minimum energy E_i^{m}. Here, unlike in Table III, for each model only three sets of results, obtained by MCM and by FMCM at $T_{\mathrm{MC}} = 280$ and 400 K, are provided. As in Table III (see discussion in the previous paragraph), the populations found with both methods for the first few bins are similar (results boldfaced). For the higher energy bins (up to $\sim$ 6 kcal/mol above the GEM) MCM finds a significantly larger number of structures than FMCM at 280 K, while the latter procedure is more efficient than FMCM at 400 K. This trend is reversed for the higher energy bins, where first FMCM (280 K) and then FMCM (400 K) become more efficient than MCM (see the boldfaced numbers in the table). We found that in bins with energies larger than 8 kcal/mol above the GEM, FMCM (400 K) becomes the most efficient procedure among the three (results not shown); this is in accord with the capability of FMCM to generate high energy structures demonstrated in Table III.

SELECTION OF SEEDS FOR MC MICROSTATES

In Tables III and IV we classified the localized microstates according to the harmonic free energy and the energy, respectively. As already discussed, an important question within the framework of our methodology for analyzing NMR data of flexible molecules is which classification leads more effectively to the most stable MC microstates; i.e., should one select localized microstates as seeds for MC simulations from an energy range of, say, 2 kcal/mol above the GEM, or define the seeds according to the harmonic free energy of microstates from the same range above the GFM? Table V reveals that localized microstates of the latter category can have energy (E_i^{m}) of up to $\sim$ 6.5 kcal/mol above the GEM. This means that the contribution of the harmonic entropy to the harmonic free energy is comparable to that of the energy. An MC microstate that is initiated from such a seed is expected to have relatively high average energy, because it would span a region in conformational space with energy that for most conformations is not smaller than the seed energy. However, such an MC microstate can still be relatively stable if its entropy is large, which means that it is correlated with the harmonic entropy of the seed. To examine this point we selected three energy minimized structures of model I ($\omega = 180°$) with backbone motifs that differ significantly from the four motifs studied in ref. 5, and with energies of 2.98, 4.86, and 6.53 kcal/mol above the GEM (i.e., all above the 2 kcal/mol range). On the other hand, the corresponding values of F_i^{har} (280 K) are 1.14, 1.55, and 1.83 kcal/mol above GFM; i.e., are all located within the 2 kcal/mol range above the GFM. These structures were taken to be seeds for metropolis MC simulations (carried out in the same way as described in ref. 5), which yielded three samples (i.e., MC microstates) of 79,200 conformations; their free energies were calculated by the local states (LS) method.[27]

In the lower part of Table VI results are presented for these MC microstates (6, 7, and 8). In the upper part we provide, for comparison, results from table V of ref. 5, obtained for the GEM and the GFM seeds (1 and 2); i.e., the seed that led to the most stable MC microstate (3), and two additional seeds with higher energy (4 and 5). For the sake of uniformity, all the differences Δ are calculated with respect to structure 3. The table reveals that the differences in both the harmonic and the MC entropies are comparable to the corresponding differences in the energy, as pointed out in the previous paragraph; therefore, in this case the energy alone does not constitute a suitable criterion for stability, and including the entropy contribution (through the free energy) is mandatory.

The results are presented in an increasing order of the seed energies E_i^{seed}, that, as expected, are always smaller than the corresponding average MC energies, E_i^{ave}. The two sets are relatively correlated: i.e., the E_i^{ave} results also appear in an increasing order for the pairs 2 and 3, and 4 and 5. Notice, however, that the energies of the first pair can be considered equal, within the statistical error, and the relative low MC energy of structure 5 probably reflects the existence of lower energy minimized structures within the conformational region covered by MC microstate 5. The results for $T\Delta S_i^{\mathrm{har}}$ and $T\Delta S_i^{\mathrm{LS}}$ are obviously less correlated

TABLE VI.
Results for Thermodynamic Properties of Localized Microstates and Corresponding MC Microstates.

i	E_i^{seed}	E_i^{ave}	ΔE_i^{seed}	$T\Delta S_i^{har}$	ΔF_i^{har}	ΔE_i^{ave}	$T\Delta S_i^{LS}$	ΔF_i^{LS}	p_i^{LS}
1	−9.70	−2.67	−0.56	−0.65	0.09	−0.23	−1.08	0.85	0.03
2	−9.46	−2.44	−0.32	0.38	−0.70	0.04	−0.01	0.05	0.14
3	−9.15	−2.48	0.00	0.00	0.00	0.00	0.00	0.00	0.15
4	−7.90	−1.24	1.25	0.15	1.10	1.24	−0.81	2.05	0.00
5	−7.77	−2.27	1.38	−0.10	1.48	0.21	0.11	0.10	0.13
6	−6.72	0.40	2.43	1.98	0.45	2.88	1.93	0.95	
7	−4.84	0.81	4.30	3.45	0.85	3.29	1.84	1.45	
8	−3.17	2.33	5.98	4.84	1.14	4.81	3.06	1.75	

Entries 1–5 were taken from table V of ref. 5; entries 6–8 are new. For convenience we kept the notations of ref. 5. The differences Δ of the various properties are calculated with respect to structure 3 that was found to have the most stable MC microstate. The results are presented in an increasing order of the seed energies E_i^{seed}. S_i^{har} and F_i^{har} are the harmonic entropy and free energy, respectively, of the localized microstates i. E_i^{ave}, S_i^{LS}, and F_i^{LS} are the average energy, entropy, and free energy of MC microstate i; the latter two quantities were obtained with the LS method.[27] All these results are in kcal/mol. p_i^{LS} is the population of MC microstate i.[5] The statistical errors of E_i^{ave} are smaller than ± 0.02. The errors of ΔF_i^{LS} are ± 0.15 kcal/mol, which lead to errors of up to ± 0.03 in p_i^{LS}. The latter errors are correlated due to the normalization condition of the probabilities.[5]

than those for the energy, even if the relatively large statistical errors of ΔS_i^{LS} are taken into account. All these finding were expected. First, note that the energy has a global characteristic: it depends on the 3-D structure. Therefore, the existence of a low energy minimized structure (the seed) suggests that other low energy minimized structures also exist within the conformational range spanned by the MC microstate, and therefore the average energy of the latter is also relatively low. On the other hand, the harmonic entropy reflects the local conformational fluctuations around an energy minimized structure; therefore, it has a limited ability to predict large fluctuations associated with the corresponding MC microstate. This is especially demonstrated by structure 5 that has relatively small harmonic entropy, and therefore relatively large harmonic free energy that is 2.18 kcal/mol above the GFM (structure 2); its corresponding MC microstate is characterized by relatively high entropy and low free energy.

However, for structures 6, 7, and 8, with the relatively high seed and MC energies, the harmonic and LS free energies are not only fully correlated, but, due to strong entropy contributions, are all located within the 2 kcal/mol range above the corresponding GFMs. Thus, the results for ΔF_i^{LS} are comparable to those obtained for several of the MC microstates appearing in table V of ref. 5; in particular, the result for structure 6 is only slightly larger than that of structure 1 (with a population of 0.03), and therefore should be taken into account. Structures 7 and 8 would lead to populations smaller than 0.01 and can be ignored.

The results outlined above seem to slightly prefer the energy classification of localized microstates over the harmonic free energy classification, because the former (based on a 2 kcal/mol range above the GEM) would miss structure 6, which is only moderately stable; the latter (based on a 2 kcal/mol range above the GFM) would miss structure 5, which is highly stable ($p_i^{LS} = 0.13$). On the other hand, if the range is increased to 2.18 kcal/mol, structure 5 would also be included with the free energy classification, but structure 6 would not be included with the energy classification. Thus, the results of Table VI are too limited to provide a conclusive answer to the question of efficiency. It should be pointed out, however, that one cannot rule out the existence of localized microstates with relatively large E_i^m and S_i^{har} (i.e., small F_i^{har}) that lead to highly stable MC microstates. This requires further study.

In summary, we studied the efficiency of the MCM and FMCM procedures. As expected, altogether the former was found to be a more efficient method for locating low energy minimized structures, while the latter is a better tool for generating localized microstates with low harmonic free energy that are characterized by larger structural diversity. As in ref. 4, we demonstrated that the energy and entropy of a microstate are correlated; i.e., on average localized and MC microstates with higher energy also have larger entropy. Thus, at $T = 280$ K the contributions of the energy and entropy to the free energy are comparable, which means that the energy cannot replace the free energy as a criterion of stability. We carried out

MC simulations, in addition to those of ref. 5, in an attempt to determine which type of localized microstates—those with low energy or low harmonic free energy—leads to the most stable MC microstates. Our results do not provide a conclusive answer to this question.

Acknowledgments

We are grateful to Drs. M. Vásquez and D. R. Ripoll, and Professor W. Braun for valuable discussions. We acknowledge support from the Florida State University Supercomputer Computations Research Institute, which is partially funded by the U.S. Department of Energy (DOE) under Contract DE-FC05-85ER250000. This work was also partially supported by DOE Grant DE-FG05-95ER62070.

References

1. H. A. Scheraga, In *Reviews in Computational Chemistry*, Vol. 3, K. D. Lipkowitz and D. B. Boyd, Eds, VCH Publishers, New York, 1992, p. 73.
2. M. Karplus and G. A. Petsko, *Nature*, **347**, 631 (1990).
3. M. Vásquez, E. Meirovitch, and H. Meirovitch, *J. Phys. Chem.*, **98**, 9380 (1994).
4. H. Meirovitch, E. Meirovitch, and J. Lee, *J. Phys. Chem.*, **99**, 4847 (1995).
5. E. Meirovitch and H. Meirovitch, *Biopolymers*, **38**, 69 (1996).
6. (a) H. Kessler, C. Griesinger, J. Lautz, A. Muller, W. F. van Gunsteren, and H. J. C. Berendsen, *J. Am. Chem. Soc.*, **110**, 3393 (1988); (b) D. F. Mierke, R. M. Scheek, and H. Kessler, *Biopolymers*, **34**, 559 (1994); (c) J. Yang and T. F. Havel, *J. Biomol. NMR*, **3**, 355 (1993); (d) R. Brüschweiler, M. Blackledge, and R. R. Ernst, *J. Biomol. NMR*, **1**, 3 (1991); (e) M. J. Blackledge, R. Brüschweiler, C. Griesinger, J. M. Schmidt, P. Xu, and R. R. Ernst, *Biochemistry*, **32**, 10960 (1993); (f) G. V. Nikiforovich, O. Prakash, C. A. Gehrig, and V. J. Hruby, *J. Am. Chem. Soc.*, **115**, 3399 (1993); (g) C. Landis and V. S. Allured, *J. Am. Chem. Soc.*, **113**, 9493 (1991); (h) D. O. Cicero, G. Barbato, and R. Bazzo, *J. Am. Chem. Soc.*, **117**, 1027 (1995).
7. (a) M. Saunders, K. N. Houk, Y. D. Wu, W. C. Still, M. Lipton, G. Chang, and W. C. Guida, *J. Am. Chem. Soc.*, **112**, 1419 (1990); (b) N. Weinberg and S. Wolfe, *J. Am. Chem. Soc.*, **116**, 9860 (1994); (c) R. A. Abagyan and P. Argos, *J. Mol. Biol.*, **225**, 519 (1992).
8. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
9. (a) B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, **31**, 459 (1959); (b) J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature*, **267**, 585 (1977).
10. (a) Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84**, 6611 (1987); (b) *J. Mol. Struct.* (*Theochem.*), **179**, 333 (1988).
11. (a) M. Saunders, *J. Am. Chem. Soc.*, **109**, 3150 (1987); (b) D. M. Ferguson and D. J. Raber, *J. Am. Chem. Soc.*, **111**, 4371 (1989); (c) G. Chang, W. C. Guida, and W. C. Still, *J. Am. Chem. Soc.*, **111**, 4379 (1989); (d) H. Meirovitch, *J. Phys. A*, **15**, L735 (1982); (e) H. Meirovitch, M. Vásquez, and H. A. Scheraga, *J. Chem. Phys.*, **92**, 1248 (1990).
12. (a) U. H. E. Hansmann and Y. Okamoto, *J. Comp. Chem.*, **14**, 1333 (1993); (b) B. A. Berg and T. Neuhaus, *Phys. Lett. B*, **267**, 249 (1991).
13. (a) S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. Schmidt, *Tetrahedron Lett.*, **29**, 4373 (1988); (b) S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science*, **220**, 671 (1983).
14. R. C. Schaik, W. F. van Gunsteren, and H. J. C. Berendsen, *J. Comput.-Aided Mol. Design*, **6**, 97 (1992).
15. M. Vásquez, G. Némethy, and H. A. Scheraga, *Chem. Rev.*, **94**, 2183 (1994).
16. (a) F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975); (b) G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87**, 1833 (1983); (c) M. J. Sippl, G. Némethy, and H. A. Scheraga, *J. Phys. Chem.*, **88**, 6231 (1984).
17. H. Meirovitch and E. Meirovitch, *J. Phys. Chem.*, **100**, 5123 (1996).
18. (a) F. H. Stillinger and T. A. Weber, *J. Phys. Chem.*, **87**, 2833 (1983); (b) F. H. Stillinger and T. A. Weber, *Science*, **225**, 983 (1984); (c) R. Elber and M. Karplus, *Science*, **235**, 318 (1987).
19. A. Caflisch, P. Niederer, and M. Anliker, *Proteins: Struct. Funct. Genet.*, **14**, 102 (1992).
20. A. Nayeem, J. Vila, and H. A. Scheraga, *J. Compt. Chem.*, **12**, 594 (1991).
21. B. von Freyberg and W. Braun, *J. Comp. Chem.*, **12**, 1065 (1991).
22. D. R. Ripoll and H. A. Scheraga, *J. Protein Chem.*, **8**, 263 (1989).
23. (a) N. Gō and H. A. Scheraga, *J. Chem. Phys.*, **51**, 4751 (1969); (b) K. D. Gibson and H. A. Scheraga, *Physiol. Chem. Phys.*, **1**, 109 (1969); (c) N. Gō and H. A. Scheraga, *Macromolecules*, **9**, 535 (1976).
24. A. T. Hagler, P. S. Stern, R. Sharon, J. M. Becker, and F. Naider, *J. Am. Chem. Soc.*, **101**, 6842 (1979).
25. (a) M. Karplus and J. N. Kushick, *Macromolecules*, **14**, 325 (1981); (b) D. A. Case, *Curr. Opin. Struct. Biol.*, **4**, 285 (1994).
26. (a) H. Meirovitch, E. Meirovitch, A. G. Michel, and M. Vásquez, *J. Phys. Chem.*, **98**, 6241 (1994); (b) H. Meirovitch and H. A. Scheraga, *Macromolecules*, **14**, 1250 (1981).
27. (a) H. Meirovitch, *Chem. Phys. Lett.*, **21**, 389 (1977); (b) H. Meirovitch, M. Vásquez, and H. A. Scheraga, *Biopolymers*, **26**, 651 (1987); (c) H. Meirovitch, D. H. Kitson, and A. T. Hagler, *J. Am. Chem. Soc.*, **114**, 5386 (1992); (d) H. Meirovitch, S. C. Koerber, J. Rivier, and A. T. Hagler, *Biopolymers*, **34**, 815 (1994).
28. (a) T. Schaumann, W. Braun, and K. Wüthrich, *Biopolymers*, **29**, 679 (1990); (b) B. von Freyberg and W. Braun, *J. Comp. Chem.*, **14**, 510 (1993); (c) B. von Freyberg, T. J. Richmond, and W. Braun, *J. Mol. Biol.*, **233**, 275 (1993); (d) B. von Freyberg, T. Schaumann, and W. Braun, *FANTOM User's Manual and Instructions*, ETH Zürich, Zürich, 1993.